On the maintenance of allozyme and inversion polymorphisms in Drosophila melanogaster

Kamping, Albert

Chapter 7

Summarizing discussion

Most animal and plant taxa possess high levels of genetic variation. Within species, often considerable differentiation for various genetically determined traits exists between populations from different geographic origins. A major question in evolutionary genetics is whether these differences in genetic constitution reflect adaptation to particular environmental conditions, are merely due to chance processes or have a historic background. When natural selection is responsible for genetic differentiation it can be asked at which level natural selection occurs. As many genetically variable traits show similar correlations with environmental gradients, selection on a particular polymorphism can lead to genetic changes in functionally and/or genetically linked traits.

In this thesis the results of a study on the adaptive value and evolution of some genetic polymorphisms in *Drosophila melanogaster* are presented. It is tried to causally relate genetic constitution with environmental conditions. In this respect, the functional and chromosomal interactions of two protein coding genes, alcohol dehydrogenase (*Adh*) and α-glycerophosphate dehydrogenase (*αGpdh*), and a chromosomal inversion [*In(2L)t*], are investigated. The three polymorphisms are all located on the left arm of the second chromosome. *In(2L)t* includes the *αGpdh* gene, while the *Adh* gene is located just outside this inversion. Almost all natural populations of *D. melanogaster* contain two common electrophoretically distinguishable alleles both for *Adh* (Adh^S and Adh^F) and for *αGpdh* ($\alpha Gpdh^S$ and $\alpha Gpdh^F$). *In(2L)t* is nearly always associated with the Adh^S and the $\alpha Gpdh^F$ alleles, while *Standard* (*ST*) chromosomes carry all possible combinations of alleles, including $Adh^S/\alpha Gpdh^F$. The frequency of Adh^S increases from around 0.10 in populations from temperate climates to above 0.90 in tropical regions. $\alpha Gpdh^F$ frequencies range from about 0.50 in temperate regions to above 0.90 in the tropics. *In(2L)t* frequencies range from less than 0.05 in populations from temperate climates to around 0.50, with extremes up to 0.70, in the tropics (references in Van Delden and Kamping 1989, 1997; Kamping and Van Delden 1999a; Van 't Land *et al.* 2000).

Genetic fluctuations in a seminatural population of *D. melanogaster*

Adh and *αGpdh* allele frequencies strongly fluctuated over a period of 25 years (i.e. about 400 generations) in a seminatural population of *D. melanogaster* kept in a tropical greenhouse (**chapter 2**). Many short-term and a few long-term fluctuations occurred for both genes in this period. The pattern of the fluctuations was not random: comparison of allele frequencies between seasons showed significantly higher Adh^S and $\alpha Gpdh^F$ frequencies in summers compared to winters. This systematic difference among seasons is clear evidence for the action of natural selection: i.e. differences in fitness between genotypes in relation with environmental differences among seasons. Apart from their significant positive correlation with environmental temperature, Adh^S and $\alpha Gpdh^F$ allele frequencies are positively correlated. The observed non-random association of alleles (gametic disequilibrium) of the two genes in many samples could indicate epistatic interaction between the two loci. Gametic

disequilibrium is positively correlated with $Adh^S/\alpha Gpdh^F$ haplotype frequency and also with environmental temperature. As both allele combinations $Adh^S/\alpha Gpdh^F$ and $Adh^F/\alpha Gpdh^S$ are in excess in the case of increase as well as decrease of Adh^S and $\alpha Gpdh^F$ allele frequencies, it seems unlikely that selection on the allozyme polymorphisms is the main cause for the observed changes. The chromosomal inversion *In(2L)t*, which is associated with the Adh^S and $\alpha Gpdh^F$ allele combination, appears to be largely responsible for fluctuations in allele frequencies at the *Adh* and *αGpdh* loci. In fact, the allozyme loci are hitchhiking with *In(2L)t*. Also deviations from the expected *Adh* genotypic ratios in some samples can be explained by selection at the chromosome level. It is convincingly shown that natural selection is operating on *In(2L)t* polymorphism, which has a large impact on the allozyme polymorphisms by chromosomal association.

Changes in environmental temperature will influence various factors, like egg production and larval density. Therefore, intraspecific competition for food and differences in developmental rate among genotypes and karyotypes are probably important fitness characters in natural *D. melanogaster* populations. Such temperature-related characters may be responsible for the observed fluctuations in genetic composition of the greenhouse population. No sharp distinction can be made between functional and chromosomal associations of the *Adh* and *αGpdh* loci as cause for the non-random association of alleles observed in this population. Moreover, non-random association of alleles observed in natural *D. melanogaster* populations may also be caused by several other factors like genetic drift, subdivided populations, migration and non-random mating.

Genetic changes in *D. melanogaster* populations with different genetic constitution

A logical next step was to separate the effects of environmental stress on each of the polymorphisms and to establish how interactions through functional and/or chromosomal linkage occur (**chapter 3** and **chapter 4**). For this purpose experimental populations with different allozyme and karyotype constitution were constructed, and the genetic composition was followed under various environmental conditions in the course of time. The *In(2L)t* polymorphism was studied in a genetic background in which either none, one or both allozyme loci were polymorphic. On the other hand, each allozyme polymorphism was studied in the presence or absence of variation at the other allozyme locus and in the presence or absence of inversion polymorphism. This allowed the analysis of multigenic reactions on basis of monogenic reactions. **Chapter 3** deals with effects of various rearing temperatures on the allozyme and *In(2L)t* polymorphisms while in **chapter 4** effects of population density, development time and environmental ethanol are presented. The environmental conditions were chosen because of their presumed relevance with respect to natural habitats of *D. melanogaster*. The allele and karyotype frequency data clearly show that natural selection is acting on each of the polymorphisms, but the direction and magnitude highly depends on genetic background of the other polymorphisms and on environmental conditions. The most relevant findings will be summarized.

In(2L)t frequencies

An overall strong disadvantage of *In(2L)t* was observed in experimental populations at lower temperatures, at high larval densities with 14 days generation interval (HD14), and at food supplemented with ethanol. At 29.5°, 33° and at high larval densities with 21 days generation interval (HD21), *In(2L)t* frequencies were highest but not significantly deviating from the initial frequency of 0.50. Frequencies in the latter environments agree with average *In(2L)t* frequencies in populations from tropical regions (references in Van Delden and Kamping 1991; Van 't Land 1997; Veuille *et al.* 1998; Van 't Land *et al.* 2000). The observed *In(2L)t* frequencies in the various environments are concordant with the higher survival rates at high temperature and the longer development time of *In(2L)t* homokaryotypes (Van Delden and Kamping 1989, 1991). *In(2L)t* apparently contains genetic variants for slower juvenile development, and compared with control conditions at 25°, the specific allelic content provides selective advantage at high temperature and disadvantage at low temperature and ethanol-rich environments. At extremely high temperature, significant superiority of *In(2L)t/ST* heterokaryotypes was observed: the advantageous effect in *In(2L)t* homokaryotypes is probably counteracted by genes with deleterious effects or by a general effect of homozygosity of the *In(2L)t* region (about 15% of the genome) as *In(2L)t* arrangements are expected to be genetically uniform to a large extent (**chapter 6**).

Adh and αGpdh allele frequencies

In monomorphic *ST* populations kept at different temperatures, larger departures from the initial allele frequency of 0.50 were observed for both *Adh* and *αGpdh* in populations segregating for one of the loci, compared with populations segregating for both loci. However, significant differences among temperatures were absent within each of these groups. The larger departure of Adh^S from the initial frequency of 0.50 compared with $\alpha Gpdh^S$, agrees with lower Adh^S than $\alpha Gpdh^S$ frequencies in *D. melanogaster* populations from regions with a temperate climate (Parkash and Shamina 1994; Bubli *et al.*1996; Van 't Land 1997).

Individuals carrying the $\alpha Gpdh^F$ allele showed faster development under high-density conditions. This agrees with frequency fluctuations observed in the seminatural population (**chapter 2**; Kamping and Van Delden 1999a) and experimental data by Marinkovic *et al.* (1987). High-density conditions may resemble the tropical situation with high numbers of individuals and high $\alpha Gpdh^F$ frequencies, whereas at higher latitudes, with slower development, individuals carrying the $\alpha Gpdh^S$ allele probably have a relative advantage as compared to the tropical situation. Populations from temperate regions will have periods with low population sizes because of food shortage and unfavourable weather conditions. Genetic drift and migration will be important in those populations. These extreme environmental conditions are generally associated with a decrease in metabolic rate, which implies that selection may occur at the level of energy carriers. The higher flight ability of $\alpha Gpdh^{SS}$ genotypes, through a better energy supply of the $\alpha GPDH^{SS}$ allozyme to flight muscles at low temperature (Barnes and Laurie-Ahlberg 1986), may be relevant in this respect.

The higher Adh^F and $\alpha Gpdh^S$ allele frequencies on food supplemented with ethanol compared to control food agrees with earlier findings (Cavener and Clegg 1981; Van Delden

and Kamping 1989). Differentiation in *Adh* and *αGpdh* allele frequencies between *D. melanogaster* populations from wineries and their surroundings (e.g. Hickey and McLean 1980; Alonso-Moraga and Muñoz-Serrano 1986) may be explained (at least partly) by the strongly reduced fitness of *In(2L)t* karyotypes in the presence of ethanol, which means that selective effects on the allozyme loci could have been overestimated by ignoring *In(2L)t* polymorphism.

The large differentiation of *Adh* as well as *αGpdh* allele frequencies due to different density conditions and the presence of ethanol compared to the effect of various temperatures shows that the role of the former environmental factors is more important for genetic differentiation at the *Adh* and *αGpdh* loci than the effects of various temperatures with optimal food conditions (Table 1).

Table 1. Summary of the changes in *In(2L)t*, Adh^S and $\alpha Gpdh^F$ frequencies in experimental populations with initial frequencies of 0.50.

Environment	***In(2L)t***	Adh^S	$\alpha Gpdh^F$
20°	—	0	0
25° (Control)	—	0	0
29.5°	0*	0	0
33°	0*	0	0
HD14	—	—	+
HD21	0	+	—
Ethanol	—	—	—

Populations were kept under optimal food conditions at four different temperatures and under high larval density conditions with transfer times of 14 and 21 days (HD 14 and HD 21), and food supplemented with ethanol at 25°. For *In(2L)t*, the *Adh* and *αGpdh* backgrounds are combined. For *Adh* and *αGpdh* only *Standard* chromosomes are involved, while the *αGpdh* respectively *Adh* backgrounds are combined in this overview. — : significant decrease in frequency; 0 : minor changes in frequency; + : significant increase in frequency; * : overdominance.

Interactions between the polymorphisms

The experimental populations polymorphic for both *Adh* and *αGpdh*, either polymorphic for *In(2L)t* and *ST* or monomorphic for *ST*, were started with maximum gametic disequilibrium between the *Adh* and *αGpdh* loci. Initially only $Adh^S/\alpha Gpdh^F$ and $Adh^F/\alpha Gpdh^S$ gametes were present. The other two gametes ($Adh^F/\alpha Gpdh^F$ and $Adh^S/\alpha Gpdh^S$) appeared with time, depending on the number of generations, recombination fraction between the *Adh* and *αGpdh* loci and fitness differences between *Adh*/*αGpdh* genotypes. In experimental populations varying for all three polymorphisms, the initial gametic disequilibrium between *Adh* and

αGpdh alleles was maintained under all environmental conditions (**chapter 3** and **4**) due to the strongly suppressed recombination in *In(2L)t/ST* heterokaryotypes. The consequence of the high and stable gametic disequilibrium is a correlated response in *Adh* and *αGpdh* allele frequencies. The two allozyme genes and the *In(2L)t* polymorphism genetically behave as one gene, with three genotypes. Homozygotes and heterozygotes are homozygous or heterozygous, respectively, for all three polymorphisms and probably for many other genes as *In(2L)t* and *ST* are genetically differentiated (**chapter 6**). Changes in allele frequencies at the *Adh* locus in this population type are associated with a change at the *αGpdh* locus and *vice versa*, and are in fact governed by changes in *In(2L)t* frequencies. Observed gametic disequilibria among the *Adh* and *αGpdh* loci in this population type are due to hitchhiking with *In(2L)t* and the magnitude of the hitchhiking effect is nearly maximal. The correlated response in *Adh* and *αGpdh* allele frequencies was also observed in the seminatural greenhouse population which was followed for many years (**chapter 2**; Kamping and Van Delden 1999a).

In *In(2L)t*-free populations, gametic disequilibrium between *Adh* and *αGpdh* decreased with number of generations. The faster decay of gametic disequilibrium at high rearing temperature compared to low rearing temperature is caused by a significantly higher recombination rate at high temperature. The observed increase in recombination rate in response to high-temperature stress is consistent with a general acceleration of evolutionary change under extreme environmental conditions (Parsons 1988). Under high-density conditions and on food supplemented with ethanol, the rate of decay of gametic equilibrium was influenced by epistatic interactions between *Adh* and *αGpdh*. Fitness interactions between *Adh*/*αGpdh* two-locus genotypes under ethanol and high-density stress are ascribed to the functional relationship of the two allozyme polymorphisms (e.g. McKechnie and Geer 1988; Oudman *et al.* 1994). Gametic equilibrium values were ultimately reached in all environments in *In(2L)t*-free populations, though after varying numbers of generations. Gametic disequilibria between these loci observed in wild *D. melanogaster* populations seem to be caused by reduced recombination frequencies, as a consequence of the presence of *In(2L)t*, rather than by epistatic interaction between the loci.

Changes in *In(2L)t* frequencies were dependent on the constitution at the *Adh* and *αGpdh* loci. In general, populations polymorphic for *Adh* had lower *In(2L)t* frequencies than monomorphic Adh^{SS} populations, while populations polymorphic for *αGpdh* exhibited higher *In(2L)t* frequencies than monomorphic $\alpha Gpdh^{FF}$ populations: the selective disadvantage for *In(2L)t* at low temperatures is reinforced by the *Adh* polymorphism and reduced by the *αGpdh* polymorphism. At high temperature, populations monomorphic for Adh^{SS} and $\alpha Gpdh^{FF}$, exhibited higher *In(2L)t* frequencies than populations polymorphic for *Adh* and/or *αGpdh*. Populations fixed or close to fixation for Adh^{S} and $\alpha Gpdh^{F}$ resemble natural tropical populations where *In(2L)t* frequencies are relatively high (e.g. Anderson *et al.* 1987; Van 't Land 1997; Van 't Land *et al.*1999; Van 't Land *et al.* 2000).

Experimental populations polymorphic for *In(2L)t* exhibited in most cases a depressing effect of *In(2L)t* on Adh^{S} and $\alpha Gpdh^{F}$ frequencies: strongest effect was observed on ethanol supplemented food and under high density with two-weeks generation interval (disadvantage for slowly developing individuals) and the effect diminished with increasing temperature

while at 33° even a reversal was observed. The negative effects of *In(2L)t* at low temperatures are stronger when *αGpdh* is polymorphic and also stronger when *Adh* is monomorphic. Differences in relative fitness of *In(2L)t* and *ST* karyotypes, depending on genetic constitution of the *Adh* and *αGpdh* genes, emphasise the role of the genetic background in selection on particular polymorphisms.

Selection pressure on the three polymorphisms

Selection coefficients calculated from the changes in *In(2L)t* frequencies and assuming selective disadvantage for *In(2L)t/In(2L)t* homokaryotypes only, yielded high values in some environments. For example, at 20°, at high larval density with 14 days generation interval (providing disadvantage for slowly developing genotypes) and at ethanol supplemented food, selection coefficients deduced from frequency changes were 0.47, 0.51 and 0.83, respectively, in experimental populations monomorphic for Adh^{SS} and $\alpha Gpdh^{FF}$. The assumption that selection is acting exclusively against *In(2L)t/In(2L)t* homokaryotypes is based on differences in developmental time between karyotypes in these environments (Van Delden and Kamping 1991; Van 't Land 1997). The two high-temperature environments are characterised by a significant excess of heterokaryotypes in many samples, indicating overdominance under these conditions. The observed excess of heterokaryotypes agrees with higher fitness values of heterokaryotypes for some fitness traits under high-temperature conditions (Van Delden and Kamping 1991, 1997; Kamping and Van Delden 1999b).

The much higher differentiation among environments for the *In(2L)t* polymorphism as compared to the two allozyme polymorphisms points to stronger selection pressures on *In(2L)t* in extreme environments. Selection coefficients calculated from frequency changes in the experimental populations and assuming selection against one of the homozygotes, do not exceed 0.05 for the *Adh* and *αGpdh* polymorphisms associated with *ST* arrangements (except for *Adh* on ethanol medium, where the selection coefficient against Adh^{SS} is 0.18), while selection coefficients for *In(2L)t/In(2L)t* homokaryotypes reach up to ten fold higher values in extreme environments. Therefore, genetic divergence among, and temporal variation within natural *D. melanogaster* populations seems to be more influenced by selection on chromosomal variation than on allozyme variation. The ratio of $Adh^{S}/\alpha Gpdh^{F}$ haplotypes associated with *In(2L)t* and *ST* will determine the magnitude of fluctuations in *Adh* and *αGpdh* allele frequencies in natural populations.

Fitness differences between karyotypes under high-temperature conditions

The causal relation between the *In(2L)t* polymorphism and high-temperature resistance was studied for some fitness components in **chapter 5**. Sterility in both sexes was induced by exposing *In(2L)t* and *ST* homo- and heterokaryotype juveniles as well as adults to 33° for some time. Fertility can be restored after a recovery period at 25°. Fertility restoration in *In(2L)t* homo- and heterokaryotypes was significantly higher than in *ST* homokaryotypes for both sexes and heterokaryotype superiority was positively correlated with severity of the high-temperature stress. A genetic increase in fertility restoration was observed for all karyotypes after 10 generations selection at 33° (including a recovery period at 25° before starting a new generation), indicating the presence of genetic variation for fertility restoration.

increasing differences in relative fitness between homokaryotypes with increasing latitude (Fig. 1). In case of heterokaryotype superiority, and assuming different ratios of the selection coefficients against the two homokaryotypes, *In(2L)t* will reach equilibrium frequencies as shown in Figure 1. These equilibrium frequencies are only dependent on the ratio of the selection coefficients and cannot provide any information on the intensity of selection. However, the number of generations needed to reach the equilibrium frequency will increase with lower absolute values of the selection coefficients.

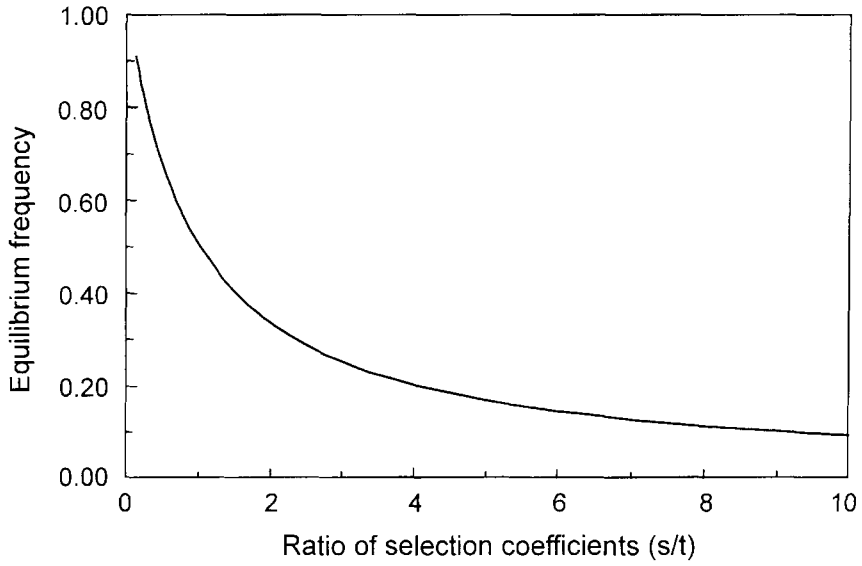

Figure 1. The relation between the ratio of selection coefficients (*s* and *t*) against the two homokaryotypes and *In(2L)t* equilibrium frequencies under the overdominance model. Fitnesses of *In(2L)t/In(2L)t*, *In(2L)t/ST* and *ST/ST* are : 1-*s*, 1 and 1-*t*, respectively.

Data concerning the relationship between allozyme heterozygosity at single loci and fitness or fitness-related characters in optimal and stressful environments are scarce. However, it is often assumed that allozyme polymorphisms are maintained by a higher overall fitness of the heterozygotes (references in Van Delden 1982 and Chambers 1988). Without disturbing effects of *In(2L)t*, an excess of heterozygotes for both single and combined allozyme loci was observed at high temperature and under high-density conditions in more samples than expected. This indicates a role for balancing selection directly acting on the allozyme loci in maintaining these polymorphisms. As shown in **chapter 3** and **4**, the relative fitnesses of *Adh* as well as *αGpdh* genotypes also depend on genetic constitution of the other locus. The observed positive relationship between severity of environmental conditions and excess of heterozygotes at the *Adh* and *αGpdh* loci, however, is mainly explained by associative overdominance through the hitchhiking effect with *In(2L)t* (**chapter 2**, **3** and **4**).

With respect to the *Adh*, *αGpdh* and *In(2L)t* polymorphisms fifteen multiple genotypes can be distinguished (Table 2). Four of these genotypes are *In(2L)t/ST* heterokaryotypes: all with $Adh^S/\alpha Gpdh^F$ associated with the *In(2L)t* chromosome, but each with a different *Adh/αGpdh* allele combination (*FF*, *FS*, *SF* or *SS*) on the *ST* chromosome. The four *In(2L)t/ST* genotypes may exhibit different fitnesses, through interactions between the allozyme and *In(2L)t* polymorphisms, and the fitness relations may differ among environmental conditions.

Table 2. *Adh/αGpdh* allele combinations for *Standard* and *In(2L)t* chromosomes and the corresponding genotypes.

	Standard				***In(2L)t***
Haplotypes	***FF***	***FS***	***SF***	***SS***	***SFI***
FF	*FF*	*FH*	*HF*	*HH*	*HF ST/I*
FS		*FS*	*HH*	*HS*	*HH ST/I*
SF			*SF*	*SH*	*SF ST/I*
SS				*SS*	*SH ST/I*
SFI					*SF I/I*

The first and second letter of the haplotypes indicate the *Adh* and *αGpdh* allele, respectively; *I*: *In(2L)t* chromosome; no chromosome indication means *Standard* chromosome. Genotypes for the *Adh* and *αGpdh* polymorphisms are indicated by *F*, *S* or *H*, where *F* and *S* indicate homozygote *Fast* and *Slow*, respectively, and *H* stands for heterozygote. *ST/I*: heterokaryotype *Standard/In(2L)t*; *I/I*: homokaryotype *In(2L)t/In(2L)t*; no chromosome indication: homokaryotype *Standard/Standard*.

Some relevant examples of such heterokaryotype superiority with varying fitness relations between heterokaryotypes with different *Adh/αGpdh* genotypes are given in Table 3. All *ST* and *In(2L)t* homozygotes are assumed to have similar fitnesses. These examples lead to "equilibrium" frequencies which are rather unusual for natural populations, but show that the impact on the allozyme polymorphisms depends on fitness relation between heterokaryotypes with different *Adh/αGpdh* genotypes. The equilibrium frequencies which are the result of heterokaryotype superiority (Table 3A and 3B) are in fact the starting frequencies of the experimental populations which are followed for their genetic composition under different environmental conditions (**chapter 3** and **4**). Further investigation learned that superiority of the two *In(2L)t/ST* heterokaryotypes which are heterozygous for *Adh* (Table 3C, example 3) leads to a complete association of Adh^S and *In(2L)t* and maximum gametic disequilibrium between *Adh* and *αGpdh* as observed in *D. melanogaster* populations from Mediterranean climates. Superiority of the two *In(2L)t/ST* heterokaryotypes which are heterozygous for *αGpdh* (Table 3C, example 4) leads to complete association of $\alpha Gpdh^F$ and *In(2L)t*, but is never found in natural *D. melanogaster* populations.

In order to explain the influence of *In(2L)t* polymorphism on the geographic distribution pattern of the *Adh* and *αGpdh* polymorphisms in tropical, Mediterranean and temperate regions, simulations were performed by assuming fitness differences between *In(2L)t* and *ST* homo- and heterokaryotypes only (Table 4). It should be noted that the *Adh* and *αGpdh* allele frequencies found in tropical regions and Mediterranean climates can be explained fully by overdominance of *In(2L)t/ST* and assuming varying fitness relationships between heterokaryotypes with different *Adh/αGpdh* constitution.

Table 3. Fitness differences among *In(2L)t/ST* heterokaryotypes with different *Adh/αGpdh* genotypes and its consequences for the *Adh* and *αGpdh* polymorphisms (see text for further explanation).

Fitness values *In(2L)t/ST*				**Equilibrium frequencies**			
SF	***SH***	***HF***	***HH***	Adh^S	$\alpha Gpdh^F$	***In(2L)t***	D/D_{max}
A:							
1.2	1.0	1.0	1.0	1.00	1.00	0.50	-
1.0	1.2	1.0	1.0	1.00	0.50	0.50	-
1.0	1.0	1.2	1.0	0.50	1.00	0.50	-
1.0	1.0	1.0	1.2	0.50	0.50	0.50	1.00
B:							
1.0	1.2	1.2	1.2	0.50	0.50	0.50	1.00
1.2	1.0	1.2	1.2	0.50	1.00	0.50	-
1.2	1.2	1.0	1.2	1.00	0.50	0.50	-
1.2	1.2	1.2	1.0	1.00	1.00	0.50	-
C:							
1.0	1.0	1.0	1.0	0.50	0.50	0.00	0.00
1.2	1.2	1.2	1.2	0.75	0.75	0.50	0.34
1.0	1.0	1.2	1.2	0.50	0.75	0.50	1.00
1.0	1.2	1.0	1.2	0.75	0.50	0.50	1.00

Various fitness relations between the four *In(2L)t/ST* genotypes (see Table 2) are shown, with the corresponding equilibrium frequencies for *Adh*, *αGpdh* and *In(2L)t*, and gametic disequilibrium values (D/D_{max}) between *Adh* and *αGpdh*. Fitness values for all other genotypes are adjusted to 1. Initial frequencies of the data presented: Adh^S = 0.50; $\alpha Gpdh^F$ = 0.50; *In(2L)t* < 0.01. A: One *In(2L)t/ST* genotype with arbitrarily chosen higher fitness value than all other 14 genotypes (see Table 2). B: Fitness of one *In(2L)t/ST* genotype is similar to fitness of all *In(2L)t* and *ST* homokaryotypes but lower than the other three *In(2L)t/ST* genotypes. C: situation for no heterokaryotype superiority; heterokaryotype superiority similar for all *In(2L)t/ST* genotypes; heterokaryotype superiority for *In(2L)t/ST* heterozygous for *Adh* and for *αGpdh*, respectively.

Table 4. Theoretical fitness values of *ST/ST*, *In(2L)t/In(2L)t* and the four *In(2L)t/ST* genotypes leading to equilibrium frequencies for *Adh*, *αGpdh* and *In(2L)t* and gametic disequilibrium values (D/D_{max}) between the allozyme loci as observed in populations from different geographic regions. Simulations (1000 generations) were performed by assuming random mating, infinite population size and no recombination in *In(2L)t/ST* heterokaryotypes. *s/t* : ratio of selection coefficients against *In(2L)t* (*s*) and *ST* (*t*) homokaryotypes (see Fig. 1).

Fitness values *In(2L)t/ST*					**Equilibrium frequencies**			
SF	***SH***	***HF***	***HH***	***s/t***	Adh^S	$\alpha Gpdh^F$	***In(2L)t***	D/D_{max}
Tropical regions:								
1.18	1.20	1.20	1.00	1.00	0.90	0.90	0.50	0.01
1.18	1.20	1.20	1.10	1.00	0.87	0.87	0.50	0.01
1.15	1.20	1.20	1.00	1.00	0.84	0.84	0.50	0.01
1.15	1.20	1.20	1.10	1.00	0.80	0.80	0.50	0.09
Mediterranean regions:								
1.00	1.00	1.10	1.10	3.00	0.25	0.63	0.25	1.00
1.00	1.00	1.10	1.10	4.00	0.20	0.60	0.20	1.00
1.00	1.00	1.10	1.10	5.00	0.17	0.58	0.17	1.00
1.00	1.00	1.10	1.10	6.00	0.14	0.57	0.14	1.00
Temperate regions:								
1.00	1.00	1.04	1.04	12	0.18	0.54	0.08	0.38
1.00	1.00	1.04	1.04	16	0.22	0.53	0.05	0.20
1.00	1.00	1.04	1.04	20	0.27	0.52	0.04	0.13
1.00	1.00	1.04	1.04	24	0.30	0.51	0.03	0.09

The time needed to reach the equilibrium frequencies depends on the level of overdominance and takes on average 250 generations for the examples in the tropics and 450 generations for the Mediterranean climates. After simulation for 1000 generations, equilibrium frequencies were not reached in the case of the fitness relations presented for temperate climates (based on the low *In(2L)t* frequencies in this climate), though for the upper two examples the *Adh* and *αGpdh* allele frequencies and the association level between Adh^S and *In(2L)t* were comparable with those from natural populations. This suggests that *D. melanogaster* populations from temperate regions may not be in equilibrium. An other possibility is that selection intensity on the allozyme polymorphisms is higher in temperate climates than in tropical and Mediterranean regions. This is supported by the stronger effects of density (restricted food conditions) and ethanol (**chapter 4**) on the *Adh* and *αGpdh* polymorphisms compared to effects of various temperatures (**chapter 3**).

The results point to a large effect of *In(2L)t* on the maintenance of the geographic distribution and on the seasonal fluctuations of the *Adh* and *αGpdh* allele frequencies. The magnitude of this effect highly depends on the fraction of $Adh^S/\alpha Gpdh^F$ haplotypes associated

with *In(2L)t* and on the fitness relations between heterokaryotypes with different *Adh/αGpdh* haplotypes on the *ST* chromosome. Moreover, the genetic content of *In(2L)t* chromosomes may vary among geographic regions, resulting in different interactions with the allozyme polymorphisms.

DNA variation in *In(2L)t* and *ST* chromosome arrangements

In(2L)t inversions are expected to have originated from a unique mutational event and to possess lower levels of variation than *ST* chromosomes due to the strongly reduced recombination in *In(2L)t/ST* heterokaryotypes. This hypothesis was studied with three different DNA-based techniques as the resolving power of allozyme electrophoresis was insufficient (**chapter 6**). DNA variation in *In(2L)t* and *ST* chromosomes from a Dutch (Groningen) and a French (Vernet, Pyrenees Orientales) population was surveyed at three levels, with the following techniques: 1) the whole *In(2L)t* region (about 13000 kb) was assayed with the Random Amplified Polymorphic DNA (RAPD) technique, 2) two gene regions associated with *In(2L)t* were analysed for Restriction Fragment Length Polymorphisms (RFLPs): *Adh* [close to the proximal *In(2L)t* breakpoint] and *αGpdh* [located in the central part of *In(2L)t*], 3) part of the *αGpdh* gene (0.6 kb) including exon and intron regions was analysed by means of DNA sequencing.

A summary of genetic differentiation between *In(2L)t* and *ST* as revealed by the different DNA techniques is presented in Table 5.

Table 5. Differentiation within and between *In(2L)t* and *ST* chromosomes from a Dutch and a southern French *D. melanogaster* population, assessed with different molecular techniques.

Technique	Within populations		Between populations	
	Groningen	Vernet	*In(2L)t*	*Standard*
A				
RAPD	+	+	+	+
RFLP *Adh*	-	-	-	-
RFLP *αGpdh*	+	-	+	-
sequencing *αGpdh*	+	-	+	-
B				
allozymes	-	-	-	-
RAPD	-	-	+	+

A. Within the *In(2L)t* region (RAPD : whole *In(2L)t* region, about 13000 kb; RFLP : the *Adh* gene region [2.7 kb probe] close to the proximal *In(2L)t* breakpoint, and the *αGpdh* gene region [6.0 kb probe] located in the central part of the inversion; sequencing : 0.6 kb of the *αGpdh* structural gene). **B.** Outside the *In(2L)t* region (allozymes and RAPD : random sampling of the rest of the genome). + : significant differentiation; - : no differentiation.

In addition, allozyme loci located outside the *In(2L)t* region (N=17; 11 polymorphic) and randomly distributed over the rest of the genome, showed no geographic differentiation within and between *ST* and *In(2L)t*. This indicates that for allozymes the genomic backgrounds for *ST* and *In(2L)t* are not differentiated both at the intra- and the inter-population level.

DNA variation in the whole In(2L)t region

Although the RAPD technique detects polymorphisms throughout the genome, it was possible with specially designed Experimental Lines (ELs) to localise RAPD markers associated with the *In(2L)t* region. The construction of these ELs was performed in such a way that the genomic backgrounds of *In(2L)t* and *ST* were similar in each of the ELs. Thus, the ELs are polymorphic for the *In(2L)t* region only (i.e. about 15% of the genome), and *ST* and *In(2L)t* homo- and heterokaryotypes are present in each of these lines. The three karyotypes can be identified by different *Adh*/*αGpdh* genotypes: $Adh^{SS}/\alpha Gpdh^{FF}$ double homozygotes are homozygous for *In(2L)t*, double homozygotes $Adh^{FF}/\alpha Gpdh^{SS}$ are homozygous for *ST* and double heterozygotes for *Adh* and *αGpdh* are heterozygous for *In(2L)t* and *ST* (Fig. 2). Only 15 out of 63 observed polymorphic DNA fragments detected in the *In(2L)t* region were unique for a karyotype, whereas 26 were unique for a locality and 22 were absent in one locality and unique for a karyotype in the other locality. This indicates that *In(2L)t* harbours a surprisingly high level of RAPD polymorphisms. Geographic differentiation for *In(2L)t*, was even higher than differentiation between *In(2L)t* and *ST* for each of the populations. Moreover, polymorphisms shared by the two chromosome types were observed, indicating at least some genetic exchange between *In(2L)t* and *ST*.

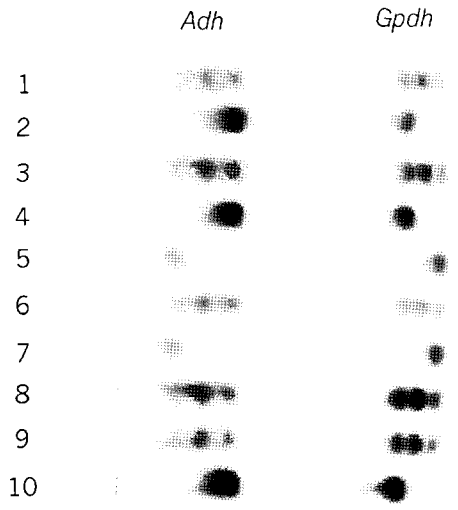

Figure 2. *Adh* and *αGpdh* allozyme banding patterns of individuals from a specially designed experimental line. Left zone of bands: *Adh*; right zone: *αGpdh* (see text). Lanes 1, 3, 6, 8 and 9: $Adh^{SF}/\alpha Gpdh^{SF}$ double heterozygotes (*In(2L)t/ST* heterokaryotypes); lanes 2, 4 and 10: $Adh^{FF}/\alpha Gpdh^{SS}$ homozygotes (*ST* homokaryotype); Lanes 5 and 7: $Adh^{SS}/\alpha Gpdh^{FF}$ homozygotes (*In(2L)t* homokaryotypes).

DNA variation in the Adh and αGpdh gene regions

DNA polymorphisms, detected with the RAPD technique, are located throughout the whole *In(2L)t* region. Analyses of the *Adh* region (close to the proximal *In(2L)t* breakpoint and reflecting the evolutionary history of the inversion) and the *αGpdh* region (located in the central part of the inversion) supply information concerning levels of DNA variation and differentiation within and between *In(2L)t* and *ST* in different parts of the *In(2L)t* region.

DNA variation in the *Adh* region was measured by the use of RFLP analysis. Nucleotide heterozygosity for *In(2L)t* and *ST* was estimated as $2.01 * 10^{-3}$and $8.20 * 10^{-3}$, respectively. Thus, two randomly chosen *In(2L)t* chromosomes differ for one out of 500 nucleotides, while two *ST* chromosomes differ by one out of 122 nucleotides in the *Adh* region. A higher level of nucleotide variation for *ST* compared to *In(2L)t* has been observed for both populations, while geographic differentiation was absent for each of the chromosome arrangements.

RFLP as well as sequence data of the *αGpdh* region showed a low level of variation for *In(2L)t* from the Groningen population, while the level of variation for *In(2L)t* from the Vernet population was similar to *ST*. The latter arrangement showed similar levels of nucleotide variation for within and between population comparisons. Polymorphisms shared by the two chromosome types were found in the Vernet population. Nucleotide heterozygosity in the *αGpdh* region was estimated as $12.63*10^{-3}$ for *In(2L)t* and as $7.50*10^{-3}$ for *ST*. Thus, the overall (populations combined) level of DNA variation in the *αGpdh* region is even higher for *In(2L)t* than for *ST*.

In(2L)t showed a six-fold higher level of DNA variation for the *αGpdh* region than for the *Adh* region (two-tailed Wilcoxon signed rank test; $P < 0.01$), while *ST* showed similar levels of DNA variation for the two gene regions. Genetic exchange within the *In(2L)t* region, either between two *In(2L)t* regions or between *In(2L)t* and *ST* by means of the occurrence of double crossovers and/or gene conversion (Rozas and Aguadé 1994; Popadic *et al.* 1995), is expected to happen more frequently with increasing distance from the chromosomal breakpoints (Ashburner 1989). Consequently, it will occur more frequently in the *αGpdh* region than in the *Adh* region. The discrepancy for the *Adh* and *αGpdh* regions with respect to levels of variation among *In(2L)t* and *ST* arrangements may therefore be explained by their chromosomal positions in relation to the *In(2L)t* breakpoints, rather than by different selection pressures acting on the *Adh* and *αGpdh* polymorphisms.

Geographic variation in In(2L)t content

As inversion frequencies are generally very low in temperate climates, but much higher in Mediterranean climates and tropical regions, genetic exchange among *In(2L)t* chromosomes and between *In(2L)t* and *ST* chromosomes will occur more frequently in the southern French population than in the Groningen population. This agrees with the relatively high level of DNA variation in the *αGpdh* region for *In(2L)t* from Vernet, the presence of shared polymorphisms and the absence of differentiation between *In(2L)t* and *ST* in this population.

The different levels of DNA variation in the *αGpdh* region for *In(2L)t* from the two populations and the distinction in differentiation between *In(2L)t* and *ST*, may be ascribed to stochastic processes like founder effects and/or bottlenecks in effective number of inversion

karyotypes. In this respect, observed DNA variation in the *αGpdh* region associated with *In(2L)t*, is positively correlated with *In(2L)t* frequencies in the populations. In addition, *In(2L)t* derived from the Groningen population, contains six unique nucleotide positions in a 605 bp *αGpdh* region. This can partly be explained by the maintenance of spontaneous recessive mutations (even those which are deleterious) in heterokaryotypes by heterokaryotype superiority (**chapter 3, 4** and **5**), leading to fixation in the inversion by chance. Moreover, levels of genetic variation within and between *In(2L)t* and *ST* arrangements from different geographic areas probably depend on the historical distribution pattern of the species (Veuille *et al.* 1998).

The results indicate that the genetic content of *In(2L)t* varies among geographic regions and therefore may vary with latitude. The latitudinal distribution in *In(2L)t* frequencies may then partly be caused by genetic content co-varying with latitude, e.g. as observed for insertion-deletion polymorphism in the *Adh* gene (Berry and Kreitman 1993). Consequently, fitness differences between individuals carrying *In(2L)t* and/or *ST* chromosomes may vary with geographic origin. Fitness components or fitness-related characters such as high-temperature resistance, developmental rate and body weight which show consistent fitness differences between karyotypes from different geographic origins (Van Delden and Kamping 1991, 1997; Van 't Land 1997; Kamping and Van Delden 1999b), are expected to be (more frequently) controlled by genes located in breakpoint regions.

Synopsis of DNA variation in In(2L)t and ST

Although cytologically identical natural inversions may not always be monophyletic in origin (Caccone *et al.* 1998), the observed results do not contravene a unique origin of *In(2L)t*: the *Adh* region (close to the proximal breakpoint, reflecting the evolutionary history) shows a four-fold lower level of nucleotide heterozygosity for *In(2L)t* than for *ST* and a low level of genetic exchange between *In(2L)t* and *ST*. In view of the lack of differentiation within *In(2L)t* as well as between *In(2L)t* and *ST* in this region, the inversion *In(2L)t* probably has a relatively recent origin.

In(2L)t chromosomes exhibit a considerable level of DNA variation and geographic differentiation as detected with RAPD analysis of the whole *In(2L)t* region and DNA analysis of the *αGpdh* region. The difference in levels of genetic variation between *In(2L)t* and *ST* depends on the geographic origin but also on the chromosomal position and nature of the DNA stretches under investigation. The higher level of differentiation obtained with the RAPD technique compared to RFLP and sequencing of two functional genes can be explained by "sampling" different types of DNA. RAPD fragments mainly consist of non-functional (neutral) DNA sequences, while the two allozyme genes contain functional and selectively constrained DNA sequences with a low rate of evolution.

Despite the substantial genetic exchange between *In(2L)t* and *ST* in the *αGpdh* region, *In(2L)t* is exclusively associated with the $\alpha Gpdh^F$ allele. The low number of silent nucleotide polymorphisms observed around the allozyme determining site of *αGpdh* in *In(2L)t* chromosomes can be explained by directional selection on that site (favouring the *F* allele) leading to reduction of nucleotide variation at neutral linked sites by hitchhiking (Aguadé *et al.* 1989; Charlesworth *et al.* 1993). Thus the pattern of polymorphic nucleotide sites

indicates that the association between *In(2L)t* and $\alpha Gpdh^F$ is maintained by natural selection. Particular allelic combinations in the *In(2L)t* region which are of adaptive significance apparently are maintained by purifying selection.

Concluding remarks

The study of three genetic polymorphisms simultaneously under highly variable natural conditions and various genetically and environmentally controlled experimental conditions, enables the analysis of interactions between these polymorphisms. The results point to very complex interactions in natural *D. melanogaster* populations, depending on allele and karyotype frequencies, on environmental conditions and on population demography. Studies on genetic variation of a single genetic polymorphism can therefore easily lead to wrong conclusions about the mechanisms underlying the maintenance and evolution of that polymorphism. By ignoring *In(2L)t* polymorphism e.g. effects of natural selection on the *Adh* and/or *αGpdh* polymorphism can be misinterpreted (mostly overestimated), as *In(2L)t* is always associated with the $Adh^S/\alpha Gpdh^F$ allele combination. It is clear that hitchhiking effects of the allozyme polymorphisms will increase with increasing fractions of $Adh^S/\alpha Gpdh^F$ haplotypes associated with *In(2L)t*. In natural *D. melanogaster* populations from Mediterranean climates, (almost) all Adh^S alleles are associated with *In(2L)t* (e.g. Van 't Land *et al.* 2000). Frequency fluctuations at the *Adh* locus in those climates can therefore be explained by selection on the *In(2L)t* polymorphism with passive hitchhiking of the *Adh* polymorphism. This probably explains the instability in Adh^S frequencies in Mediterranean climates as described by David *et al.* (1989).

The large fitness differences between *In(2L)t* and *ST* karyotypes (a gene region containing about 2000 genes) which are observed under particular environmental conditions can be further explored by identifying the genes that affect fitness. The observed karyotype diagnostic DNA markers can be used as a starting point for the localisation and identification of genes. Since the completion of the *D. melanogaster* genome project in March 2000, the DNA sequences and chromosomal locations of all genes in the *In(2L)t* region are known. This information can be used for measuring differences in gene expression, if any, between *In(2L)t* and *ST* karyotypes under particular environmental conditions. The micro-array procedure allows measurements of expression of many genes simultaneously under various experimental conditions. This molecular approach looks promising in detecting genes with different levels of expression between *In(2L)t* and *ST* karyotypes. Both population genetic and molecular genetic research will be necessary to link these data with fitness differences at the individual as well as at the population level. It will be a big challenge for future research to characterise and localise that part of the 2000 genes in the *In(2L)t* region which affects fitness and to investigate how epistatic interactions and/or pleiotropic effects between those genes contribute to fitness differences between karyotypes.